# INVERSE RANDOM SAMPLING (STROKE

- Group Members:
- Jaafar Bin Farooq 18k-1294
- Muhammad Irham Rahim

18k-0150

Muhammad Hasan18k-0294





According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths.

This dataset is used to predict whether a patient is likely to get stroke

### RETRIEVING DATA

Dataset was picked from https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset

95 % did not have stroke 5% had stroke

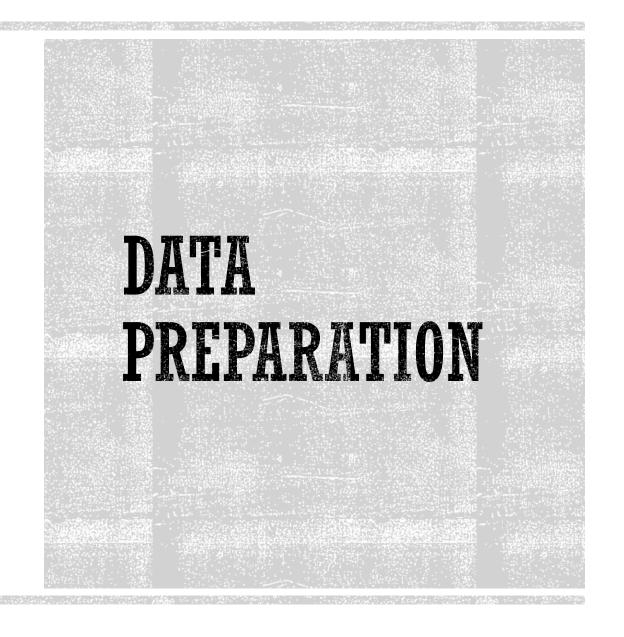
#### 11 features 1 Target Variable

#### Attribute Information

- 1) id: unique identifier
- 2) gender: "Male", "Female" or "Other"
- 3) age: age of the patient
- 4) hypertension: 0 if the patient doesn't have hypertension, 1 if the patient has hypertension
- 5) heart\_disease: 0 if the patient doesn't have any heart diseases, 1 if the patient has a heart disease
- 6) ever\_married: "No" or "Yes"
- 7) work\_type: "children", "Govt\_jov", "Never\_worked", "Private" or "Self-employed"
- 8) Residence\_type: "Rural" or "Urban"
- 9) avg\_glucose\_level: average glucose level in blood
- 10) bmi: body mass index
- 11) smoking\_status: "formerly smoked", "never smoked", "smokes" or "Unknown"\*
- 12) stroke: 1 if the patient had a stroke or 0 if not
- \*Note: "Unknown" in smoking\_status means that the information is unavailable for this patient



- Replaced null values in BMI with mean
- Replaced missing values in smoking\_status using KNNImputer
- Feature Selection based on correlation
- One hot encoding on categorical variables



## DATA EXPLORATION

Count plot was used for discrete features

#### Distribution and Box plot was used for continuous/Numeric features

- Being unmarried reduces your risk of a stroke
- Being a smoker or a formerly smoker increases your risk of having a stroke
- more than 25% of stroke cases They had hypertension
- Female and male both have equal number of stroke cases while there is not any single case of stroke in other gender type.
- Patient with private job have more number stroke cases then patient who are self employed or have a government job
- Stroke has the highest correlation with age.
- Patients with stroke having higher avg\_glucose\_level





#### DATA MODELING

- IRUS Algorithm was applied
- 24 base classifier with 4 different Algorithms
- Classification Algorithms used:
  - 1. DecisionTreeClassifier
  - 2. RandomForestClassifier
  - 3. MLPClassifier
  - 4. LogisticRegression





## **CONCLUSION**

 By using IRUS we were able to classify the minority class better

